

# АДАПТАЦИЯ ПАКЕТОВ КВАНТОВОЙ ХИМИИ ДЛЯ РАСЧЕТОВ НА GPU ЭЛЕКТРОННОЙ СТРУКТУРЫ ТВЕРДОТЕЛЬНЫХ СИСТЕМ

В.Е. Гусаков<sup>1)</sup>, В.И. Белько<sup>2)</sup>, Н.Н. Дорожкин<sup>2)</sup>

<sup>1)</sup>ГНПО Научно-практический центр НАН Беларуси по материаловедению,  
ул. П. Бровки, 19, Минск, 220072, Беларусь, gusakov@ifftp.bas-net.by

<sup>2)</sup>Белорусский государственный университет,  
пр. Независимости, 4, Минск, 220030, Беларусь, belko@bsu.by, dorozhkin@bsu.by

В настоящей работе рассматривается адаптация пакетов квантовой химии Quantum Espresso и LASTO для расчетов электронной структуры сложных твердотельных систем, в частности, защитных покрытий на основе нитридов переходных металлов, с использованием графических процессоров серии GeForce и технологии NVIDIA CUDA.

## Введение

В последние годы использование графических процессоров (GPU) для научных расчетов находит все более широкое применение. Для вычислений на GPU компанией NVIDIA разработана специальная технология программирования CUDA – расширение языка программирования C. Для решения сложных научных задач обычно применяются вычислительные кластеры, содержащие тысячи и даже десятки тысяч CPU и сотни GPU. Разумеется, вычисления на таких кластерах являются весьма дорогостоящими. При этом многие интересные с научной и практической точки зрения задачи можно успешно решать и на современных многоядерных персональных компьютерах с процессорами фирм Intel и AMD и относительно недорогими видеокартами GeForce серий GTX 600-900.

Следует, однако, отметить, что прирост производительности при использовании GPU зависит от ряда факторов, которые следует принимать во внимание: 1) задача должна хорошо распараллеливаться на большое количество потоков, а взаимодействие между потоками должно отсутствовать или быть слабым; 2) одни и те же инструкции применяются к огромным массивам данных; 3) в алгоритмах желательно избегать множественного ветвления и сложных для обработки операций: деление, возведение в отрицательную степень, вычисления математических функций т.д. Именно такие вычисления характерны для программ квантовой химии.

## Основная часть

В настоящей работе выполнена адаптация пакета Quantum Espresso [2] для расчетов с использованием графических процессоров и технологии CUDA, а также пакета LASTO, реализующего линейный метод присоединенных слэйтеровских орбиталей (ЛПСО). Пакет LASTO написан на языке FORTRAN-90 и является собственной разработкой.

Конфигурация системы для тестирования:

1. Системная плата ASUSH87M-PRO;
2. Процессор Intel® Core™ i7-4770K CPU 3,4 ГГц;
3. Память 16 Гб DDR3 PC6400;
4. Видеокарта NVIDIA GeForce GTX 660, 1032.5 МГц, 2047.7 МБ;
5. HDD WDC WD10003FZEX-0 1000 Гб;
6. Linux x86 64-бит, NVIDIA CUDA: 6.0, 6.5.

Адаптация пакета Quantum Espresso для данной конфигурации осуществлялась с использованием компиляторов фирмы Intel (30-дневная пробная версия).

Предварительно изучалось быстроедействие GPU на примере умножения матриц (компилятор gfortran 2014, пробная версия). Результаты показаны на рисунке 1. Эффективность применения вычислений на GPU для матриц высокой размерности очевидна.

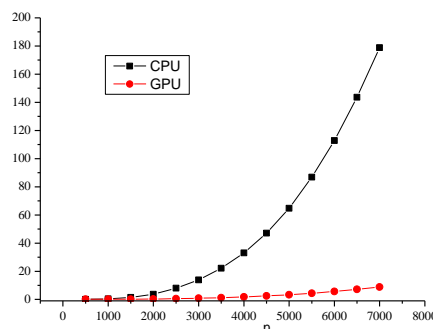


Рис. 1. Умножение матриц (двойная точность) с помощью процедуры dgemm из библиотеки BLAS; n – порядок матриц, k-точка в зоне Бриллюэна.

Возможность применения графических процессоров и технологии CUDA для высокопроизводительных вычислений в квантовой химии с помощью пакета Quantum Espresso изучена на примере самосогласованных расчетов электронной структуры твердого раствора  $Ti_{0.5}Zr_{0.5}N$ , содержащего 64 атома в элементарной ячейке. Такие твердотельные структуры широко используются в качестве защитных покрытий.

Учитывая большое количество атомов в элементарной ячейке, расчеты проводились с использованием одной k-точки (Г-точка).

Для твердого раствора  $Ti_{0.5}Zr_{0.5}N$  оценка проводилась по 40 итерациям одной петли самосогласования как при использовании только CPU, так и при вычислениях с помощью GPU. Результаты расчетов приведены на рисунках 2 и 3.

Анализ результатов расчетов позволяет сделать вывод, что оптимальным будет вариант 1 MPI+GPU, приводящий к ускорению выполняемой задачи в 1.37 раза по сравнению с одним MPI

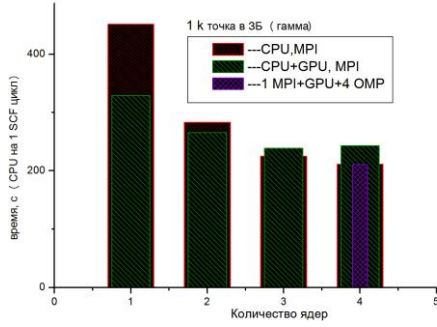


Рис. 2. Время выполнения петли самосогласования для твердого раствора  $\text{Ti}_{0.5}\text{Zr}_{0.5}\text{N}$ .

процессом. Остальные ядра следует загружать другими задачами.

В методе линейных присоединенных слэтеровских орбиталей волновая функция представляется как линейная комбинация так называемых присоединенных слэтеровских орбиталей [3, 4]:

$$\Psi_{\vec{k}}(\vec{r}) = \sum_N c_N \Psi_N(\vec{r}) \quad (1)$$

где

$$\Psi_N(\vec{r}) = \frac{1}{\sqrt{N}} \sum_{\vec{R}_v} e^{i\vec{k}\vec{R}_v} \phi_{nlm}(\vec{r} - \vec{r}_i - \vec{R}_v), \quad (2)$$

здесь индекс  $N = \{i, nlm\}$ ,  $\vec{r}_i$  — позиция  $i$ -го атома в элементарной ячейке,  $\vec{R}_v$  — вектор решетки, а функции  $\phi_{nlm}(\vec{r})$  — слэтеровские орбитали вне МТ-сфер и точные решения радиального уравнения Шредингера внутри МТ-сфер, гладко присоединенные к слэтеровским орбиталам на границе МТ-сфер:

$$\Psi_N = \begin{cases} \chi_N(\vec{r}) = \frac{1}{\sqrt{N_c}} \frac{1}{v} \sum_{\vec{g}} e^{i(\vec{k}+\vec{g})\vec{r}} \tilde{\phi}_N(\vec{k}+\vec{g}) \times \\ \quad T_i(\vec{g}), \quad \vec{r} \in \Omega_2 \\ \varphi_N(\vec{r}) = \frac{1}{\sqrt{N_c}} \frac{1}{v} e^{i\vec{k}(\vec{r}+\vec{r}_i)} \times \\ \quad \sum_{\Lambda} [\beta_{N,k\lambda} g_{k\lambda}(\vec{r}_k) + \alpha_{N,k\lambda} \dot{g}_{k\lambda}(\vec{r}_k)] Y_{\Lambda}(\vec{r}_k), \quad r \in \Omega_1 \end{cases} \quad (3)$$

Здесь  $N_c$  — число элементарных ячеек в кристалле,  $V$  — объем элементарной ячейки,  $\vec{k}$  — волновой вектор,  $\vec{g}$  — вектор обратной ячейки,  $\alpha$  и  $\beta$  — матрицы разложения решения радиального уравнения в  $i$ -й сфере по решениям в  $k$ -х сферах, которые находятся из условия непрерывности базисной волновой функции и ее производной на поверхности  $i$ -сферы;  $g_{k\lambda}$  и  $\dot{g}_{k\lambda}$  — это решение радиального уравнения и соответствующая энергетическая производная решение

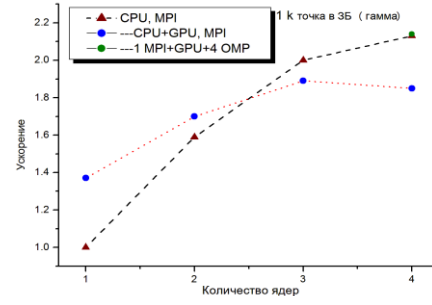


Рис. 3. Ускорение выполнения петли самосогласования для твердого раствора  $\text{Ti}_{0.5}\text{Zr}_{0.5}\text{N}$  при использовании GPU.

радиального уравнения для  $k$ -го атома и орбитального квантового числа  $\lambda$ ;  $Y_{\Lambda}$  — сферическая гармоника,  $\Lambda$  — совокупность индексов  $\{\lambda\mu\}$ ,  $\Omega_1$  — область внутри МТ-сфер или область первого типа, а  $\Omega_2$  — область вне МТ-сфер или область второго типа и структурный фактор  $T_i(\vec{g}) = e^{-i\vec{g}\cdot\vec{r}_i}$ , а  $\tilde{\phi}_N$  представляет собой трехмерный Фурье-образ слэтеровской орбитали [3]. Коэффициенты  $\alpha$  и  $\beta$ , необходимые для гладкой сшивки базисных функций на поверхности МТ-сфер, приведены в [5].

Выражение для матричных элементов оператора Гамильтона имеет следующий вид [5,6]:

$$H_{NN'} = \varepsilon \sum_{\vec{r}_k, \lambda} [\beta_{N,k\lambda}^* \beta_{N',k\lambda} + \alpha_{N,k\lambda}^* \alpha_{N',k\lambda} \langle \dot{g}_{k\lambda} | \dot{g}_{k\lambda} \rangle] + \\ + \frac{1}{v} \sum_{\vec{g}} \tilde{\phi}_N^*(\vec{g} + \vec{k}) \tilde{\phi}_{N'}^j(\vec{g} + \vec{k}) e^{i\vec{g}(\vec{r}_i - \vec{r}_j)} \left( |\vec{g} + \vec{k}|^2 \right) - \\ - \frac{1}{v^2} \sum_{\vec{g}, \vec{g}'} \tilde{\phi}_N^*(\vec{g} + \vec{k}) \tilde{\phi}_{N'}^j(\vec{g}' + \vec{k}) e^{i\vec{g}\vec{r}_i} e^{-i\vec{g}'\vec{r}_j} \times \\ \times \sum_{\vec{r}_k} ((\vec{g} + \vec{k})(\vec{g}' + \vec{k})) 4\pi R_k^2 \frac{j_1(|\vec{g}' - \vec{g}| R_k)}{|\vec{g}' - \vec{g}|} e^{i(\vec{g} - \vec{g}')\vec{r}_k} + \\ + \sum_{\Lambda \vec{r}_k} (R_{MT}^k)^2 \left[ g_{k\lambda}^* \dot{g}_{k\lambda} (\beta_{N,k\lambda}^* \alpha_{N',k\lambda} + \alpha_{N,k\lambda}^* \beta_{N',k\lambda}) + \right. \\ \left. g_{k\lambda}^* g_{k\lambda}' \beta_{N,k\lambda}^* \beta_{N',k\lambda} + \dot{g}_{k\lambda}^* \dot{g}_{k\lambda}' \alpha_{N,k\lambda}^* \alpha_{N',k\lambda} \right] \quad (4)$$

Матричные элементы матрицы перекрывания получаются в результате замены оператора Гамильтона на единичный оператор:

$$O_{NN'} = \sum_{\vec{r}_k, \lambda} [\beta_{N,k\lambda}^* \beta_{N',k\lambda} + \alpha_{N,k\lambda}^* \alpha_{N',k\lambda} \langle \dot{g}_{k\lambda} | \dot{g}_{k\lambda} \rangle] + \\ + \frac{1}{v} \sum_{\vec{g}} \tilde{\phi}_N^*(\vec{g} + \vec{k}) \tilde{\phi}_{N'}^j(\vec{g} + \vec{k}) e^{i(\vec{r}_i - \vec{r}_j)\vec{g}} - \\ - \frac{4\pi}{v^2} \sum_{\vec{g}, \vec{g}'} \tilde{\phi}_N^*(\vec{g} + \vec{k}) \tilde{\phi}_{N'}^j(\vec{g}' + \vec{k}) \times \\ \times \sum_{\vec{r}_k} e^{i\vec{g}(\vec{r}_i - \vec{r}_k)} e^{i\vec{g}'(\vec{r}_k - \vec{r}_j)} (R_{MT}^k)^2 \frac{j_1(|\vec{g}' - \vec{g}| R_{MT}^k)}{|\vec{g}' - \vec{g}|} \quad (5)$$

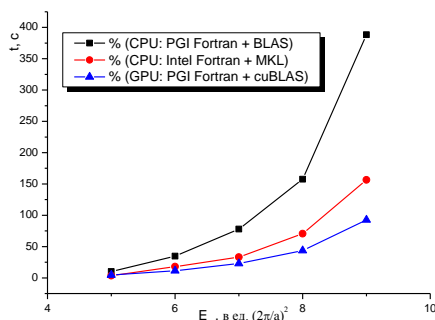


Рис. 4. Ускорение вычисления матричных элементов для структуры TiC при использовании GPU.

В методе ЛПСО основные временные затраты связаны с вычислением матричных элементов. «Узким местом» метода является необходимость вычисления двойных сумм по векторам обратной решетки (третьи слагаемые в формулах (4), (5)).

Заметно уменьшить время расчета матричных элементов можно, если переписать слагаемые, содержащие двойные суммы в формулах (4) и (5), следующим образом:

$$-\frac{1}{v^2} \sum_{\vec{g}, \vec{g}'} (\square) = \sum_{\vec{g}} \tilde{\phi}_N^{i*}(\vec{g} + \vec{k}) \sum_{\vec{g}'} H_{\vec{g}\vec{g}'}(\vec{k}) \tilde{\phi}_N^j(\vec{g}' + \vec{k})$$

$$= \sum_{\vec{g}} \tilde{\phi}_N^{i*}(\vec{g} + \vec{k}) \sum_{\vec{g}'} H_{\vec{g}\vec{g}'}(\vec{k}) \tilde{\phi}_N^j(\vec{g}' + \vec{k}), \quad (6)$$

$$-\frac{1}{v^2} \sum_{\vec{g}, \vec{g}'} (\square) = \sum_{\vec{g}} \tilde{\phi}_N^{i*}(\vec{g} + \vec{k}) \times$$

$$\dots \sum_{\vec{g}'} O_{\vec{g}\vec{g}'}(\vec{k}) \tilde{\phi}_N^j(\vec{g}' + \vec{k}) \quad (7)$$

где  $H_{\vec{g}\vec{g}'}(\vec{k})$  и  $O_{\vec{g}\vec{g}'}(\vec{k})$  – эрмитовы матрицы.

Таким образом, вычисление двойных сумм сводится к стандартным операциям линейной алгебры: умножению эрмитовых матриц на вектор и вычислению скалярного произведения векторов.

Для этих операций существует стандартная библиотека BLAS и имеются высоко оптимизированные библиотеки: MKL, ACML, GOTO, ATLAS. Использование этих библиотек позволяет увеличить скорость вычислений в 2-3 раза.

Использование GPU может дать заметное ускорение, если в вычислениях нет сложных операций, а именно такая ситуация и реализуется при перемножении матриц. Поэтому часть кода, ответственная за вычисление матричных элементов по формулам (6) и (7), была переписана с использованием технологии CUDA и библиотеки cuBLAS, позволяющей производить матричные операции.

Тестирование пакета LASTO выполнено для расширенной элементарной ячейки карбида титана, содержащего 8 атомов в ячейке. На рисунке 4 приведено время вычисления двойных сумм в матричных элементах в зависимости от энергии обрезания плоских волн  $E_{\text{cut}}$ .

## Заключение

С использованием технологии CUDA пакеты квантовой химии Quantum Espresso и LASTO адаптированы для вычислений на GPU. Применение GPU позволяет ускорить расчеты электронной структуры при использовании обоих пакетов. Особенно эффективно ускоряются расчеты электронной структуры систем с большим количеством атомов в элементарной ячейке при использовании пакета LASTO (в несколько раз).

## Список литературы

1. <http://www.nvidia.ru/object/tesla-server-gpus-ru.html>.
2. <http://www.quantum-espresso.org>.
3. Davenport J.W. // Phys. Rev. 1984. V. B26. P.2896-2904.
4. Davenport J.W., Weinert M., Watson R.E. // Phys. Rev. 1985. V. B32. P. 4876-4882.
5. Якутович Н.Г., Дорожкин Н.Н., Анищук В.М., Навыш Б.В. // ФТТ. 2000. Т. 42, вып. 11. С. 1943-1947.
6. Гурский Л.И., Дорожкин Н.Н., Якутович Н.Г., Навыш Б.В. // Весті НАН Беларусі. Сер. фіз.-мат. навук. 2001. № 3. С.62-67.

## ADAPTATION OF QUANTUM CHEMISTRY SOFTWARE FOR THE ELECTRONIC STRUCTURE CALCULATIONS ON GPU FOR SOLID-STATE SYSTEMS

V.E. Gusakov<sup>1</sup>, V.I. Belko<sup>2</sup>, N.N. Dorozhkin<sup>2</sup>

<sup>1</sup>Scientific-Practical Materials Research Centre NAS of Belarus,  
P. Brovki str., 19, Minsk, 220072, Belarus, gusakov@iftp.bas-net.by

<sup>2</sup>Belarusian State University,  
Nezavisimosti av. 4, 220030, Minsk, Belarus, belko@bsu.by, dorozhkin@bsu.by

We report on adaptation of quantum chemistry software - Quantum Espresso and LASTO - for the electronic structure calculations for the complex solid-state systems on the GeForce series GPUs using the NVIDIA CUDA technology. Specifically, protective covering based on transition metal nitrides are considered.